J. Parasitol., 96(1), 2010, pp. 49–57

COCCIDIAN MEROZOITE TRANSCRIPTOME ANALYSIS FROM *EIMERIA MAXIMA* IN COMPARISON TO *EIMERIA TENELLA* AND *EIMERIA ACERVULINA*

Ryan S. Schwarz, Raymond H. Fetterer, George H. Rosenberg*, and Katarzyna B. Miska

Animal Parasitic Diseases Laboratory, U.S. Department of Agriculture, 10300 Baltimore Avenue, Beltsville, Maryland 20705. *e-mail: schwarz.ryan@gmail.com*

ABSTRACT: With the *Eimeria* spp. populations that infect chickens used as a model for coccidian biology, we aimed to survey the transcriptome of *Eimeria maxima* and contrast it to the 2 other *Eimeria* spp. for which transcriptome data are available, i.e., *Eimeria tenella* and *Eimeria acervulina*. The asexual intracellular development stage, the merozoite, was specifically examined, and we used expressed sequence tag (EST) analysis to provide experimental evidence of transcription and a framework for understanding the merozoite stage of *E. maxima*. Of 2,680 individual ESTs obtained, 48.2% shared most significant ($E < 10^{-5}$) homology to sequences from other apicomplexan species, primarily other *Eimeria* spp. and *Toxoplasma gondii*, and 47.5% were unique. Annotation of these ESTs enabled categorization to putative biological function and revealed an emphasis on translation, cytoskeleton, metabolism, signaling, transport, and protein folding, as well as the apicomplexan specific surface antigens and micronemes. Comparative analysis of abundantly expressed transcripts from merozoites of the 3 *Eimeria* spp. revealed a novel transcript common to all 3. Sharing no significant homology to any other sequence in public databases, this transcript was predicted to encode an *Eimeria*-specific protein (ESP) with 166–178 amino acids and 58.9–65.1% interspecific identity. A predicted signal peptide was identified, consistent with the assumption that ESP is a secreted protein. These annotated ESTs from *E. maxima* merozoites provide a resource for intra- and interspecific comparative analyses that will be useful in distinguishing the unique biology of coccidian parasites in relation to the diverse phylum of Apicomplexa.

The domesticated chicken (*Gallus gallus*) is host to a group of coccidian protozoan parasites of the genus *Eimeria*, comprised of 7 recognized host-specific species. These *Eimeria* spp. are a model population of agriculturally important parasites that belong to the same phylum (Apicomplexa) as medically important members such as *Toxoplasma gondii*, *Cryptosporidium* spp., and *Plasmodium* spp. A chicken may harbor single or multispecies *Eimeria* infections, with each species varying in the region of the intestine and/or villi parasitized, the length of prepatent and sporulation time, and in the resulting pathology (Tyzzer et al., 1932; Conway and McKenzie, 2007).

Chickens become infected when they ingest a sporulated *Eimeria* sp. oocyst from which sporozoites emerge. These sporozoites migrate to, attach, and invade epithelial cells lining specific regions of the intestinal lumen. Inside the host cell, sporozoites multiply asexually, forming a schizont within which they differentiate into merozoites that erupt out of the host cell to the extracellular matrix. These merozoites must then locate and invade new epithelial cells, a process mediated by a unique glideosome complex (Herm-Götz et al., 2002; Opitz and Soldati, 2002). As apicomplexan parasites lack cilia or flagella for locomotion and do not project pseudopodia to engage in crawling motility, the glideosome is essential to the unique form of actin–mysosin-based gliding motility employed during the invasion of host cells (Hettmann et al., 2000). Many of the molecules that comprise the glideosome are secreted by micronemes and rhoptries, specialized organelles at the apical end of the merozoite. Attachment and invasion of apicomplexan parasites to host cells also involves members of a family of proteins called surface antigens (SAGs), described from *T. gondii* (reviewed in Lekutis et al., 2001). Homologues to these SAGs have been described from *E. tenella* merozoites and sporozoites and were shown to be a diverse set of proteins linked to the cell membrane via glycosylphosphatidylinositol (GPI) anchoring (Tabarés et al., 2004). Two–three additional rounds of merozoite replication and invasion result in the destruction of the epithelium lining the intestine and cause the pathology of coccidiosis disease. Merozoites ultimately differentiate into gametes, thus initiating sexual reproduction and producing oocysts that are shed out of the chicken's digestive tract.

Of all the *Eimeria* species that infect chickens, most research, particularly molecular work, has focused on *E. tenella*. However, *E. maxima* is also a species of interest for 3 reasons. First, it is highly immunogenic, eliciting a strong immune response in chickens (Rose and Long, 1962; Blake et al., 2005), thereby making this species of important consideration with regard to the efforts toward developing and optimizing coccidia vaccines (Vercruysse et al., 2004). Second, *E. maxima* has comparatively high levels of genetic variation among the repertoire of chicken *Eimeria* spp., as evidenced by DNA sequence analysis (Barta et al., 1998; Lew et al., 2003; Cantacessi et al., 2008; Schwarz et al., 2009), phenotypic variation implied by immunologic variability among strains (Fitz-Coy, 1992; Martin et al., 1997), and the lack of cross-protective immunity to *E. maxima* using xenotypic antigens (Ding et al., 2005; Song et al., 2009). For example, immunization with the use of *E. tenella* antigens is known to confer varying degrees of cross-species immunity to *E. acervulina* and *E. necatrix*, though no cross-protection was developed against *E. maxima* (Song et al., 2009). Third, *E. maxima* is globally ubiquitous, with regional strains isolated from chicken farm facilities in Europe (Schnitzler et al., 1998, 1999), North America (Martin et al., 1997), South America (Fernandez et al., 2003), Australia (Lew et al., 2003), and Asia (Li et al., 2004).

The generation of cDNA libraries provides an essential foundation from which genetic transcriptional studies can be constructed (Adams et al., 1991). Screening cDNA library clones for an expressed sequence tag (EST) analysis is a prime example, and provides data for such research goals as gene discovery, genetic mapping, and comparative analyses (Marra et al., 1998). Although EST surveys of *E. tenella* (Wan et al., 1999) and *E. acervulina* (Miska et al., 2008) have been completed, comparative analyses have focused on intraspecific stage comparisons (Ng et al., 2002; Li et al., 2003; Miska et al., 2004), leaving interspecific transcriptome

Received 17 July 2009; revised 28 August 2009; accepted 14 September 2009.

* Department of Biology, University of New Mexico, Albuquerque, New Mexico 87131.

DOI: 10.1645/GE-2253.1

comparisons of *Eimeria* spp. unexamined prior to this study. To increase our understanding of the biological function of *Eimeria* parasites at the intracellular pathogenic stage of development, we developed a nonnormalized *E. maxima* merozoite cDNA library from which we have obtained 2,680 high-quality ESTs. These ESTs have been annotated and submitted to GenBank (GO304867-GO307546), as well as made available via an online database (http://129.24.144.93/emaxmz_final), where all results obtained from sequence homology queries are accessible. These data provide the first description of the *E. maxima* merozoite transcriptome and are compared with transcriptome profiles from the 2 other common chicken coccidia species for which EST data are available, e.g., *E. acervulina* and *E. tenella*.

MATERIALS AND METHODS

Propagation and isolation of parasites

One-day-old SexSal chickens (n = 15) (Moyer's Hatcheries Inc., Quakertown, Pennsylvania) were housed in isolation at the Animal Parasitology Unit (United States Department of Agriculture [USDA] in Beltsville, Maryland) and maintained with food and water ad libidum. At 4 wk of age, the chickens were infected with ~400,000 sporulated *Eimeria maxima* 'Tysons' strain oocysts (maintained at USDA) per bird via gavage (Fetterer and Barfield, 2003).

The following merozoite collection and purification protocol is modified from Xie et al. (1990), specifically for *E. maxima*. Infected chickens were killed at 96 hr postinfection (PI) via cervical dislocation, and midintestines were excised (from the caudal end of the duodenal loop to the egg sac diverticulum) then rinsed clean with PBS (pH 7.2). The midintestines were cut open longitudinally and gently washed of any remaining fecal matter, cut into 2–3 inch pieces, and held on ice. Tissue was immersed in mucosal digestion media (HBSS [Invitrogen, Carlsbad, California], 0.50% sodium taurodeoxycholate hydrate, 97% TLC [Sigma-Aldrich, St. Louis, Missouri], 0.25% trypsin [Sigma]) prewarmed to 41 C, and merozoites were allowed to emerge for 15–22 min under gentle agitation. Merozoites were initially filtered from the media through 4 layers of cheesecloth and pelleted at 2,000 *g*/10 min followed by washing and resuspension in HBSS ×2. The final resuspension in 150–200 ml of HBSS was allowed to filter via gravity through a modified 115-ml Nalgene (Rochester, New York) 0.2-μm membrane sterilization filter unit, which had the first layer of the membrane removed to avoid clogging. This filtered out cellular debris and allowed the merozoites to migrate through the filter into the eluate. Merozoites were then pelleted at 2,000 *g*/10 min and HBSS was carefully siphoned off before the suspension was frozen at −70 C.

cDNA library construction and screening

Total RNA was isolated from the merozoites as described previously (Miska et al., 2008) with the use of TRIzol reagent (Invitrogen) and sent to RxBioSciences (Rockville, Maryland), where small (0.4–2 kb) and large (>2 kb) -fraction cDNA libraries were prepared directionally at EcoRI and XhoI restriction sites with the use of the Uni-ZAP® XR Vector Kit (Stratagene, La Jolla, California). After mass excision to pBluescript SK(-) vector, SOLR *Escherichia coli* cells were transfected, cultured, and collected for glycerol stocks. From these stocks, serial dilutions were plated on LB+ampicillin (50 μg/ml) agar plates and grown at 37 C. Plasmids from recombinant colonies were isolated in 96-well format with the use of DirectPrep® 96 Miniprep Kit (Qiagen, Valencia, California) and inserts were sequenced with BigDye® Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, Foster City, California) with the use of vector priming sites (M13). Sequences were run on a 96-capillary ABI 3730×l DNA analyzer (Applied Biosystems).

EST clustering, assembly, and annotation

Raw sequence data files were verified for quality and accuracy with the use of Sequencher™ 4.8 software (Gene Codes Corporation, Ann Arbor, Michigan) from which flanking vector regions and low-quality regions of the sequences were trimmed. Vector, chicken, and *E. coli* contaminants were identified via Basic Local Alignment Search Tool (BLAST) (Altschul et al., 1990) with the use of VecScreen and/or BLASTN, and removed from the data set. The resulting high-quality sequences were compared to one another and assembled into contigs with the use of the assemble feature of Sequencher™ 4.8 with a minimum of 70–base pair (bp) overlap and 85% sequence identity. The resulting data set was exported in FASTA format and uploaded to a database for hidden Markov model (HMM) similarity searches against multiple databases that included nonredundant (nr), dbEST, UniProt Knowledgebase (UniProtKB)/Swiss-Prot, gene ontology (GO) (Gene Ontology Consortium, 2000) and clusters of orthologous groups for eukaryotic complete genomes (KOG). The 2,680 individual high-quality ESTs were annotated based on the results of the above-mentioned similarity searches and loaded into the National Center for Biotechnology Information's (NCBI) GenBank Expressed Sequence Tag database (dbEST) under accession numbers GO304867–GO307546.

In addition, dbEST was searched (15 April 2009) for *E. tenella* merozoite stage-specific ESTs, resulting in 7,945 explicitly identified accessioned sequences. We retrieved these sequences and compared them to one another with the use of the Sequencher™ 4.8 software as described above to identify clusters representing the most abundant genes expressed by *E. tenella* merozoites for comparison to our data. This data set is available from the authors upon request.

RESULTS

Generation and clustering of ESTs from *E. maxima* merozoites

From a nonnormalized *E. maxima* merozoite cDNA library, we isolated recombinant plasmids from 2,496 individual clones from which 3,168 end-sequencing events were run, 2,496 from the 5′ end and 672 from the 3′ end. After eliminating low-quality and obvious vector/contaminated sequences, a total of 2,680 (85%) quality sequences were analyzed via BLAST queries against multiple databases, the results of which are accessible via an online database at http://129.24.144.93/emaxmz_final. These ESTs were annotated (see the following) and deposited to dbEST at GenBank (GO304867–GO307546), making *E. maxima* the second most abundantly analyzed *Eimeria* species transcriptome to date (Table I). These data provided a total of 1,838 kb of sequence, with individual usable trace lengths that averaged 643 bp (range = 84–928 bp). The 2,680 ESTs were compared to one another and assembled to generate clusters of sequences with ≥85% identity across their overlapping regions. Of these, 1,716 sequences overlapped significantly and formed 416 distinct clusters, while 964 did not overlap significantly and were classified as singletons (Table II). The length of contiguous sequence from clustered ESTs ranged from 128 to 2,409 bp and numbered from 2 to 54 overlapping sequences.

Annotation of *E. maxima* merozoite ESTs

All 964 EST singletons and 416 EST clusters were compared against sequences from the following databases for homology: nr, dbEST, UniProtKB/Swiss-Prot, KOG, and GO. In total, 52.5% of the ESTs had significant (deemed $E < 10^{-5}$) identity to previously accessioned sequences, and 47.5% had no significant match in any sequence database and were thus unique. As shown in Figure 1, 48.2% of the ESTs had highest similarity to an apicomplexan sequence, the majority of which were *Eimeria* species (55.8%), followed by the closely related taxon *T. gondii* (42.8%). A small proportion of ESTs had highest similarity to sequences from fungi (1.2%), plants (0.9%), invertebrates (0.6%), bacteria (0.6%), vertebrates (0.5%), and other nonapicomplexan protists (0.5%). Putative functional categories based on the GO

TABLE I. Quantification of transcriptome analysis across Apicomplexa.

Taxa	No. of expressed sequence tags (ESTs) *†
Eimeria tenella	34,998
Eimeria maxima	2,700‡
Eimeria acervulina	1,430
Toxoplasma gondii	129,421
Plasmodium falciparum 3D7	37,529
Plasmodium vivax	22,238
Plasmodium berghei strain ANKA	47,085
Plasmodium berghei	11,880
Plasmodium yoelii yoelii	13,925
Plasmodium yoelii	5,007
Cryptosporidium parvum	30,138
Cryptosporidium muris RN66	27,498
Neospora caninum	25,072
Neospora hughesi	1,888
Theileria annulata	17,031
Theileria parva	4,380
Theileria orientalis	3
Sarcocystis neurona	15,382
Sarcocystis falcatula	6,332
Babesia orientalis	304
Babesia bovis	7

* As per GenBank dbEST 03 April 2009.
† All developmental stages are included in tally.
‡ Of these, 2,680 are provided by this study.

database (http://www.geneontology.org) and KOG database (http://www.ncbi.nlm.nih.gov/COG/) were assigned to ESTs (Table II). Additional categories unique to the Apicomplexa were also assigned that included genes likely to be involved in functioning of the surface antigens, micronemes, rhoptries, and dense granules (as per Li et al., 2003). Although the majority of ESTs were of unknown function (60.9%), 75% of the transcripts assigned a putative function (1,047 ESTs in total) fell into 9 categories: translation (27.5%), cytoskeletal (8.9%), glycolysis (6.5%), surface antigen (6.4%), signaling (6.1%), transport (5.9%), protein folding (5.1%), metabolism (4.7%), and microneme (4.6%). The remaining 25% of transcripts assigned a putative function were spread across 21 additional categories represented less frequently in the transcriptome (Table II). As glycolysis and the tricarboxylic acid (TCA) cycle are at the center of oxidative metabolism, we designated transcripts to these processes specifically, rather than under the broad umbrella of "metabolism." As noted above, the anaerobic metabolic pathway of glycolysis was well represented in number, with 68 ESTs (Table II) and diversity, with 7 (of 10) core catalytic enzymes identified, e.g., phosphofructokinase, fructose bisphosphate aldolase, triose phosphate isomerase, glyceraldehyde 3-phosphate dehydrogenase (GAPDH), phosphoglycerate kinase, phosphoglycerate mutase, and pyruvate kinase (Supplement to Table I available upon request). Additionally, the enzyme L-lactate dehydrogenase (LDH) occurred (3 ESTs), representing the potential for conversion of pyruvate to L-lactate via metabolic fermentation. In contrast, the tricarboxylic acid (TCA) cycle was represented by only 5 ESTs (Table II); a cluster of 4 predicted to encode acetyl coenzyme A (CoA) synthetase that produces acetyl CoA to initiate the TCA cycle, and a singleton predicted to encode succinyl CoA synthetase, a downstream enzyme catalyzing the production of succinate and CoA.

Comparison of highly expressed ESTs in 3 *Eimeria* species

A comparison of the most abundantly expressed merozoite-stage transcripts, based on the number of overlapping ESTs (≥85% identity), was made between our data set for *E. maxima* to a data set we generated from *E. tenella* merozoite sequences previously archived in GenBank dbEST, plus a previously published data set of all available *E. acervulina* merozoite ESTs (Miska et al., 2008). The sequence identities for *E. acervulina* most abundant clusters from the previous analysis were reanalyzed with the use of BLAST to verify highest homology based on slightly more stringent e-values ($E < 10^{-5}$ vs. $E < 10^{-4}$) and resulted in the identification of previously unknown cluster *Ea*mzC_27 as poly(A)-binding protein 3–like (3e-14). Of 2,680 *E. maxima*, 7,945 *E. tenella*, and 1,426 *E. acervulina* merozoite ESTs analyzed, the identities of the 18 most abundant clusters that accounted for 15.97%, 20.63%, and 9.68% of all ESTs analyzed, respectively, are graphically represented in Figure 2 and tabulated by rank in Table III. Three of the most abundant ESTs were common to more than 1 species' transcriptome; the glycolytic pathway component GAPDH 1 from *E. maxima* and *E. acervulina*, the surface antigen MZ92/120 from *E. maxima* and *E. tenella*, and 1 abundantly expressed transcript of unknown identity common to all 3 species, identified from *E. maxima* as cluster *Em*mzC_0843, with similarity (~8e-60) to abundantly expressed transcript clusters from *E. tenella* (*Et*mzC_0011) and *E. acervulina* (*Ea*mzC_15). This sequence shared no significant homology to any other sequence outside of these *Eimeria* species and thus appeared to encode an *Eimeria* sp.–specific protein that ranked in abundance sixth (*E. maxima*), second (*E. acervulina*), and third (*E. tenella*) (Table III). We examined these gene sequences further by analyzing their predicted proteins, which shared 58.9% to 65.1% amino acid (aa) identity among the 3 species between the predicted start and stop codons (Fig. 3). The proteins were predicted to comprise 174 (*E. maxima*), 178 (*E. acervulina*), and 166 (*E. tenella*) aa's, with molecular weights of 18.86, 19.30, and 18.12 kDa, respectively. Signal peptides were predicted from the first 18 aa's of each sequence with the use of SignalP v3.0 (Bendtsen et al., 2004). One stress-inducible phosphoprotein (STI1) heat shock chaperonin-binding motif domain (E = 1e-12) was predicted at the core of all 3 sequences via SMART (Letunic et al., 2006). Neither potential transmembrane regions nor GPI modification sites were identified.

Predicted surface antigens (SAGs) were abundantly represented in all 3 species (Fig. 2) with the greatest variety (5) observed from *E. tenella*. Two abundant EST clusters were predicted to encode SAGs from *E. maxima*: cluster *Em*mzC_0206 with highest identity (2e-25) to *E. acervulina* antigen MZ92/120 (GenBank AAA29080) and cluster *Em*mzC_0185 with highest identity (8e-23) to the SAG4 gene described from *E. tenella* (Tabarés et al., 2004). Both of these EST clusters from *E. maxima* ranked 11th in abundance, with each comprised of 16 overlapping ESTs (Table III). Although no characterization of antigen MZ92/120 has been published to our knowledge, this same antigen shared highest similarity (2e-21) to the most abundant *E. tenella* merozoite EST cluster, comprised of 301 overlapping ESTs (Table III). From the *E. tenella* merozoite ESTs we analyzed, 4 of the characterized

TABLE II. Distribution of *Eimeria maxima* expressed sequence tags (ESTs) according to putative functional category.

Putative function	No. of EST clusters	Total no. of ESTs from clusters	No. of EST singletons	Total no. of ESTs	Percent of all ESTs	Percent of categorized
Unknown	257	925	708	1633	60.93	n/a
Translation	52	251	37	288	10.75	27.51
Cytoskeletal	16	79	14	93	3.47	8.88
Glycolysis	7	61	7	68	2.54	6.49
Surface antigen	10	63	4	67	2.50	6.40
Signaling	8	45	19	64	2.39	6.11
Transport	11	43	19	62	2.31	5.92
Protein folding	5	41	12	53	1.98	5.06
Metabolism	10	32	17	49	1.83	4.68
Microneme	9	37	11	48	1.79	4.58
RNA processing	5	12	28	40	1.49	3.82
Motility	1	32	3	35	1.31	3.34
Binding	2	24	5	29	1.08	2.77
Energy	6	19	7	26	0.97	2.48
Transcription	1	2	22	24	0.90	2.29
Proteolysis	3	9	8	17	0.63	1.62
Cell division	2	14	3	17	0.63	1.62
Protein turnover	1	2	13	15	0.56	1.43
Chaperone	1	3	8	11	0.41	1.05
Cell redox	3	7	1	8	0.30	0.76
TCA cycle	1	4	1	5	0.19	0.48
DNA replication/repair	0	0	5	5	0.19	0.48
Chromatin structure	0	0	4	4	0.15	0.38
Rhoptry	1	2	2	4	0.15	0.38
Biosynthesis	0	0	3	3	0.11	0.29
Nucleosome assembly	1	2	1	3	0.11	0.29
Nucleotidase	1	3	0	3	0.11	0.29
Endocytosis	1	2	0	2	0.07	0.19
Stress response	1	2	0	2	0.07	0.19
Apoptosis	0	0	1	1	0.04	0.10
Dense granules	0	0	1	1	0.04	0.10
Total	416	1716	964	2680	100.00	100.00

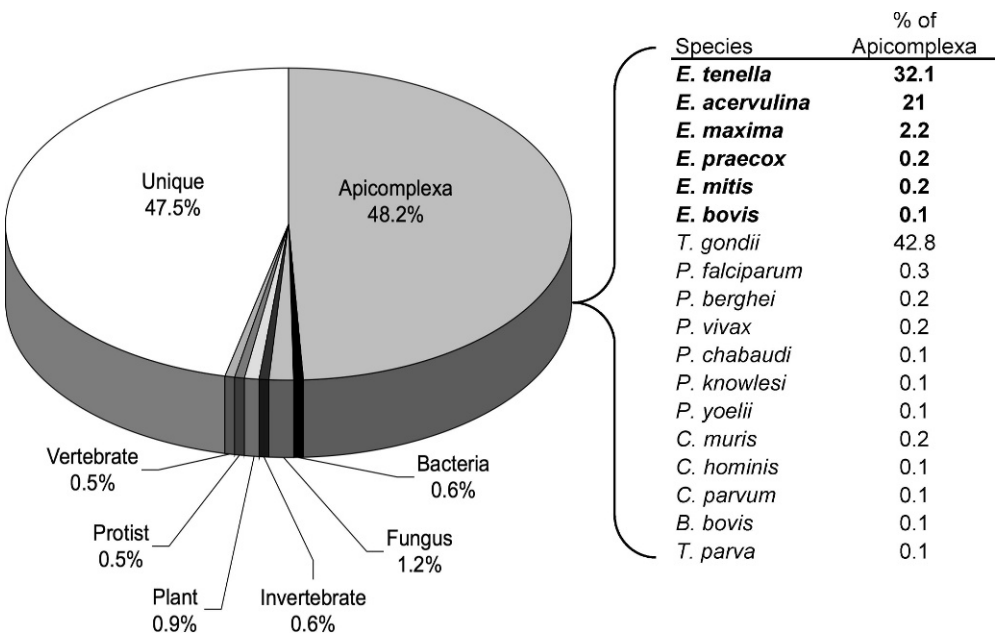
SAGs (Tabarés et al., 2004) occurred abundantly, e.g., SAG23, 2, 18, and 17 (Fig. 2), ranking 5th, 8th, 11th, and 12th in abundance, respectively. In comparison, previous analysis of *E. acervulina* merozoite ESTs (Miska et al., 2008) identified a SAG3-like cluster (4e-06) that ranked sixth in abundance and was the only abundant SAG (Table III).

Components of the glideosome, the locomotory system used specifically by apicomplexans to achieve their characteristic gliding motility, were abundantly represented in *E. maxima* merozoite ESTs with cluster *Em*mzC_0869 (n = 32) that had high identity (1e-121) to Myosin A and cluster *Em*mzC_0029 (n = 17) similar (1e-153) to inner membrane complex protein (IMC1) (Fig. 2, Table III). High expression of glideosome components was also identified in *E. tenella* merozoites as cluster *Et*mzC_0046 (n = 50) had highest similarity (6e-70) to the myosin light chain (MLC1) gene and cluster *Et*mzC_0076 (n = 234) similar (9e-43) to actin depolymerizing factor (ADP1). Though glideosome transcripts did occur from *E. acervulina* merozoites (e.g. ADP1), no representatives occurred with relative abundance.

Sequences that had no significant homology to any other sequences accessioned in the databases searched were identified as ''unique'' and included 5 abundant clusters from *E. maxima* and 1 abundant cluster each from *E. tenella* and *E. acervulina* (Fig. 2). In addition, several abundant clusters from *E. maxima* (n = 3) and *E. tenella* (n = 2) represented ''unknown'' genes that had significant similarity to uncharacterized sequences from other taxa. Combined, these unique and unknown transcript clusters represent genes abundantly expressed by *Eimeria* sp. merozoites that have yet to be described. The remaining abundant EST clusters from *E. maxima* represent genes involved in translation (ribosomal proteins and elongation factors), protein folding (heat shock proteins and cyclophilin), and signaling (phosphatidic acid phosphatase type 2A and calmodulin), which were also abundantly represented categories in the transcriptomes of the other 2 species (Wan et al., 1999; Miska et al., 2008; the present study).

DISCUSSION

Though *Eimeria* spp. are an agriculturally important group of parasites, large-scale EST analyses have been limited to *E. tenella* and *E. acervulina* with 34,998 and 1,430 ESTs currently accessioned, respectively. Here, we have provided data for a third important species, *E. maxima*, by generating a nonnormalized cDNA library from the merozoite life stage from which 2,680 annotated ESTs have been made freely available to other researchers via GenBank and an online database. This is the first transcriptome analysis from *E. maxima* and provides data that will increase the understanding of the parasite's biological



Species	% of Apicomplexa
E. tenella	**32.1**
E. acervulina	**21**
E. maxima	**2.2**
E. praecox	**0.2**
E. mitis	**0.2**
E. bovis	**0.1**
T. gondii	42.8
P. falciparum	0.3
P. berghei	0.2
P. vivax	0.2
P. chabaudi	0.1
P. knowlesi	0.1
P. yoelii	0.1
C. muris	0.2
C. hominis	0.1
C. parvum	0.1
B. bovis	0.1
T. parva	0.1

FIGURE 1. Taxonomic distribution of 2,680 ESTs from *Eimeria maxima* to currently available accessioned sequences. EST sequences that had a significant ($E < 10^{-5}$) match to a previously accessioned sequence were categorized according to the taxon to which the highest sequence identity (best match) occurred, whereas ESTs with no significant match were categorized as "unique." Matches to other *Eimeria* species are in bold type and, combined, account for the majority (55.8%) of *E. maxima* EST best matches.

functioning at this important pathogenic stage of development. These data revealed many newly identified genes, as 47.5% of ESTs shared no significant homology to currently available sequences and are thus unique (Fig. 1). This is similar to the number of novel genes identified previously from *E. acervulina* (53%) (Miska et al., 2008) and *E. tenella* (52.3%) (Wan et al., 1999) merozoite EST analyses. Despite substantial sequencing efforts on several other apicomplexans, including *T. gondii*, *Plasmodium* spp., and *Cryptosporidium* spp. (Table I), the fact that such a large proportion of *Eimeria* spp. ESTs are nonetheless unique indicate that these species have evolved highly specific transcriptomes among the Apicomplexa. This comparison emphasizes previous reports that genetic diversity in *Eimeria* spp. populations is substantial (reviewed in Beck et al., 2009) and that studies on a variety of species are necessary for comparative analyses to reveal novel adaptations in addition to conserved components.

To provide such insight, we contrasted our EST data set from *E. maxima* to the other 2 *Eimeria* spp. for which data are available. This comparative transcriptome analysis identified a novel gene abundantly expressed by all 3 *Eimeria* spp. that was predicted to encode a secreted *Eimeria* sp.–specific protein (ESP). One of the oft-lauded benefits of EST analyses from parasites is the identification of proteins conserved in the parasite, but absent from the host, making them candidates for effective intervention strategies via immunological or chemical means (Li et al., 2003; Vercruysse et al., 2004). The ESP is such a candidate, with the added benefit that it is abundantly expressed in all 3 *Eimeria* spp. From the ESP sequences we have examined here, this gene appears to have a low level of intraspecific sequence diversity (Fig. 3). Additional research into the genetic diversity, expression pattern at different developmental stages, and protein localization of ESP will be necessary to assess candidate status as a target for parasite intervention further.

From the results of our homologous gene search, we tabulated the taxa to which the best identity scores occurred. Of the ESTs that had significant identity to another accessioned sequence, *E. maxima* sequence homology was highest to other *Eimeria* spp. (55.8%) most frequently, followed closely by *T. gondii* (42.8%), leaving <2% to other taxa (Fig. 1). Although high identity to other *Eimeria* spp. was expected, high identity to *T. gondii* was also expected, as *T. gondii* is a closely related lineage to the *Eimeria* spp. clade (Kuo et al., 2008) with an abundance of accessioned sequence data. Though *Neospora* spp. are thought to be similarly close relatives to the *Eimeria* clade, these data from *E. maxima* had higher overall homology to *T. gondii* than to *Neospora* spp. However, *Neospora* spp. EST data accessioned in dbEST is approximately 1/5 of that for *T. gondii* (Table I) and may contribute to the lack of highest identity hits to *Neospora* genes.

As the cDNA library from which these ESTs were derived was not normalized, the number of ESTs should correlate with the actual abundance of transcripts present in the *E. maxima* merozoite cells. Thus, our analyses of predicted functional categorization reflect the actual contribution of the transcriptome dedicated to these categories. With 30 functional categories represented (Table II), this cDNA library contains sequences from a broad range of genes that are transcribed during the merozoite stage of development. The Apicomplexa-specific categories of surface antigens (SAGs) and microneme components were comparatively well represented in the transcriptome overall, representing 6.4% and 4.6% of all ESTs categorized. Because SAGs and microneme proteins mediate the process of host cell recognition, attachment, and invasion, this observation is

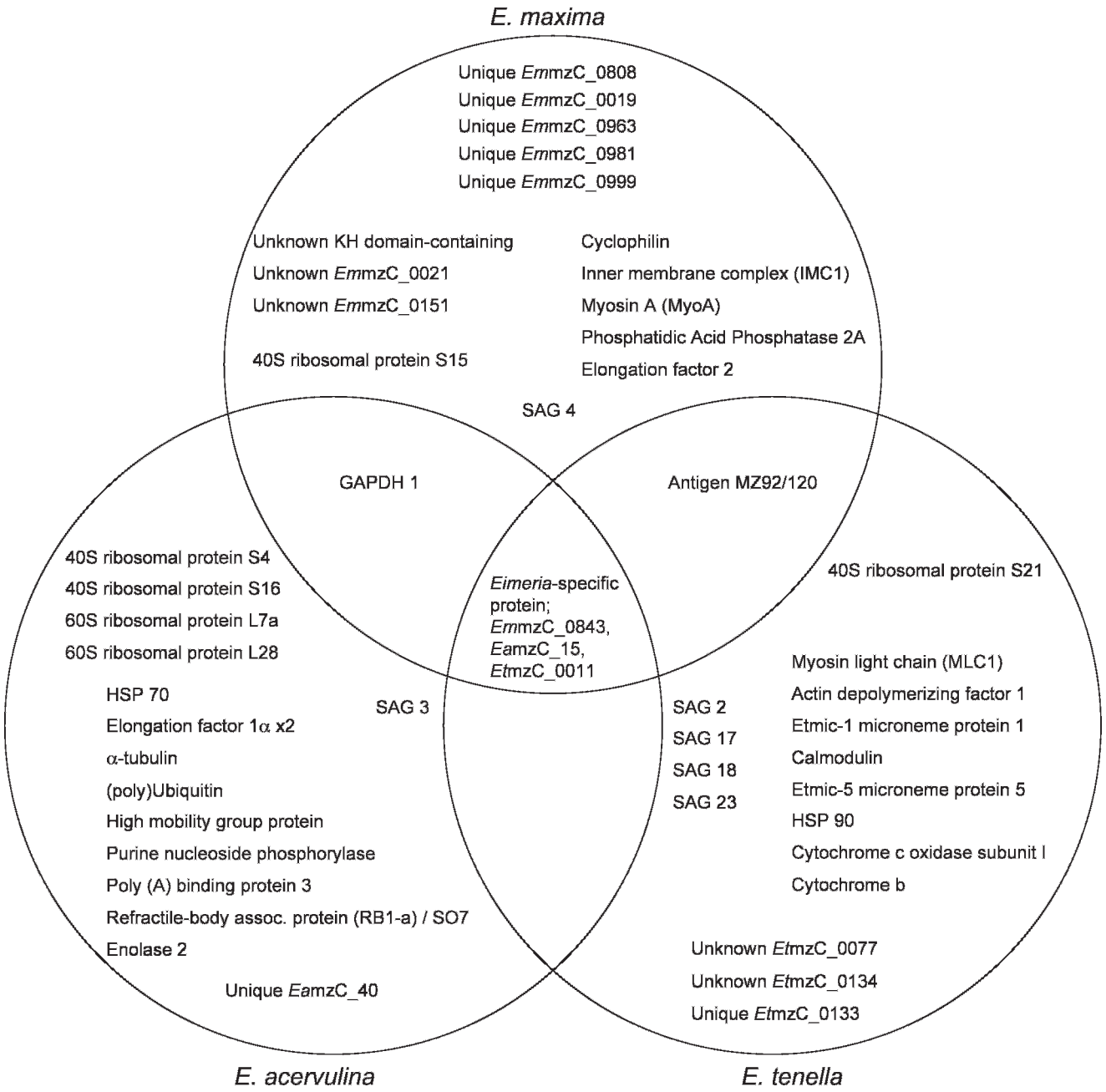


FIGURE 2. Comparison of the most abundant merozoite stage ESTs from three species of *Eimeria* hosted by the chicken. Clusters comprised of the most overlapping individual sequence reads are given within a circle for the most abundant 18 ESTs from each species. ESTs that occurred abundantly in more than 1 species fall within the overlapping regions of each species' circle. ESTs were identified by similarity ($E < 10^{-5}$) to previously identified sequences in public gene banks. Identification numbers are given for unique and unknown EST clusters, where unique sequences had no significant match to any other sequence accessioned in public databases and unknown had a significant match to an unidentified sequence.

in agreement with the importance of these processes to merozoites. The particular importance of SAG transcripts to merozoites was emphasized in our comparison of the most abundant ESTs from *E. maxima*, *E. tenella*, and *E. acervulina* (Fig. 2). From *E. maxima*, 1 EST cluster, identified as SAG4, occurred abundantly (Table III), with 5 additional SAG4 EST clusters also identified; 2 of the 5 clusters were comprised of 7 ESTs each and 3 clusters of 2 ESTs each (Supplement to Table I available upon request). That all of these SAG4-like ESTs did not overlap significantly to form a single unified cluster was due to a high level of sequence variability between each of the 6 clusters that may thus represent several members of a SAG gene family similar to that described from *E. tenella* (Tabarés et al., 2004). The diversity described from the SAG gene family of *E. tenella* revealed 2 groupings, group A (n = 12) with 28–60% aa identity and group B (n = 11) with 42–77% aa identity. As our clustering analysis was set with a cutoff of 85% nucleotide identity, similarly low levels of identity could explain the distinct clustering of SAG-like genes we identified in *E. maxima*. We also identified EST clusters with high similarity to *E. tenella* SAG7 (7e-37) and

TABLE III. Comparison of most abundant *Eimeria* merozoite stage expressed sequence tags (ESTs) by species.

Eimeria maxima merozoites		*Eimeria acervulina* merozoites*		*Eimeria tenella* merozoites†	
Gene/cluster identification‡	n§	Gene/cluster identification‡	n	Gene/cluster identification‡	n
Unique *Em*mzC_0808 [GO307275]	54	Refractile body assoc. protein (RB1-a); SO7 *Ea*mzC_9 [EH385238]	16	Antigen MZ92/120, *E. acervulina* merozoite stage (M86628) *Et*mzC_0031 [BM321883]	301
Unique *Em*mzC_0019 [GO305908]	52	Unknown *Eimeria*-specific protein; similar to *Em*mzC_0843 and *Et*mzC_0011 *Ea*mzC_15 [EH385093]	11	Actin depolymerizing factor 1 (ADP1) *Et*mzC_0076 [AI756785]	234
Unique *Em*mzC_0963 [GO307390]	36	Heat shock protein 70 *Ea*mzC_31 [EH386122]	9	Unknown *Eimeria*-specific protein; similar to *Em*C_0843 and *Ea*C_15 *EtmzC_0011 [AM948291]*	195
GAPDH 1 *Em*mzC_0033 [GO306057]	34	Unique *Ea*mzC_40 [EH386408]	9	Etmic-1 microneme protein 1 EtmzC_0094 [BM306674]	100
Myosin A *Em*mzC_0869 [GO307343]	32	Elongation factor 1α *Ea*mzC_19 [EH386417]	8	Surface antigen 23, *E. tenella Et*mzC_0083 [AI755962]	98
Unknown; *Eimeria*-specific protein; similar to *Et*mzC_0011 and *Ea*mzC_15 *Em*mzC_0843 [GO307292]	24	α-tubulin *Ea*mzC_77 [EH385952]	8	Unknown; similar to *E. tenella* clone BW1-E06; nucleoredoxin *Et*mzC_0077 [AM948455]	68
Cyclophilin-cytosolic *Em*mzC_0895 [GO307358]	23	(Poly)ubiquitin *Ea*mzC_6 [EH385398]	7	Calmodulin *Et*mzC_0273 [BM305914]	66
Unknown KH domain-containing protein *Em*mzC_0030 [GO306007]	22	60S ribosomal protein L7a *Ea*mzC_18 [EH385595]	7	Surface antigen 2, *E. tenella Et*mzC_0364 [BM305446]	65
Unknown; similar to hypothetical protein TGME49 024130 *Em*mzC_0021 [GO305960]	19	High mobility group protein *Ea*mzC_35 [EH385978]	7	Etmic-5 microneme protein 5 *Et*mzC_0074 [BM306106]	54
Inner membrane complex protein (IMC1) *Em*mzC_0029 [GO305988]	17	Purine nucleoside phosphorylase *Ea*mzC_49 [EH386199]	7	Heat shock protein 90 *Et*mzC_0002 [AM264963]	53
Antigen MZ92/120, *E. acervulina* merozoite stage (M86628) *Em*mzC_0206 [GO306495]	16	GAPDH 1 *Ea*mzC_54 [EH386063]	7	Unique *Et*mzC_0133 [BI895178]	53
Surface antigen 4, *E. tenella* CAE52296 *Em*mzC_0185 [GO306442]	16	Polyadenylate-binding protein 3; *Ea*mzC_27 [EH385086]	6	Surface antigen 18, *E. tenella Et*mzC_0105 [BE027396]	52
Unique *Em*mzC_0981 [GO307427]	16	40S ribosomal protein S4 *Ea*mzC_29 [EH385183]	6	Surface antigen 17, *E. tenella Et*mzC_0042 [AM265057]	51
Unique *Em*mzC_0999 [GO307446]	15	Surface antigen 3, *E. tenella Ea*mzC_42 [EH385401]	6	Cytochrome c oxidase subunit I *Et*mzC_0575 [BM306061]	51
Phosphatidic acid phosphatase type 2A (PAP2) *Em*mzC_0004 [GO305855]	14	60S ribosomal protein L28 *Ea*mzC_75 [EH386142]	6	Myosin light chain (MLC1) *Et*mzC_0046 [AM948402]	50
Unknown; similar to *E. tenella* S5-2 cDNA CD570411 *Em*mzC_0151 [GO306344]	14	Elongation factor 1α *Ea*mzC_94 [EH385076]	6	40S Ribosomal protein S21 *Et*mzC_0078 [BM306900]	50
Elongation factor 2 *Em*mzC_0002 [GO305837]	12	Enolase 2 *Ea*mzC_112 [EH386071]	6	Cytochrome b *Et*mzC_0080 [AI757521]	49
40S ribosomal protein S15 *Em*mzC_0032 [GO306033]	12	40S ribosomal protein S16 *Ea*mzC_127 [EH385150]	6	Unknown; similar to *Em*mzC_1009 *Et*mzC_0134 [BM306405]	49

* Results from Miska et al. (2008).
† Results obtained from analysis of ESTs mined directly from GenBank. Assembly file available upon request.
‡ Accession numbers in brackets are to a representative dbEST entry from the EST cluster.
§ n, number of merozoite stage ESTs in cluster out of 2,680 (*E. maxima*), 1,426 (*E. acervulina*), and 7,945 (*E. tenella*).

SAG11 (5e-30), with 4 and 3 overlapping ESTs, respectively, and 2 singleton ESTs similar to SAG8 (2e-07 and 7e-08) (Supplement to Table I available upon request).

This analysis of the *E. maxima* transcriptome has provided a significant database as a resource to further understanding of *Eimeria* spp. biology. For example, we presented a comparison of abundant transcripts from 3 species of *Eimeria* and found that merozoites of each species were distinct in their gene expression profiles. Although we identified similarities among the 3 species of *Eimeria* in individual genes and in functional categories abundantly expressed, quantitative gene expression analysis may be useful in corroborating and detailing the profiles of both similar and distinct transcripts from each species. Such comparative analyses among these parasites should help identify the genetic basis for their biological and pathogenic differences. This database also provides a resource that can be mined to identify variable loci via microsatellite sequences and single nucleotide polymorphisms (SNPs). These motifs can be used for the development of genetic footprints and may prove helpful in species and strain discrimination within *Eimeria* populations, as has been the case in *T. gondii* studies (Ajzenberg et al., 2002, 2004). They may also enable genetic association to specific

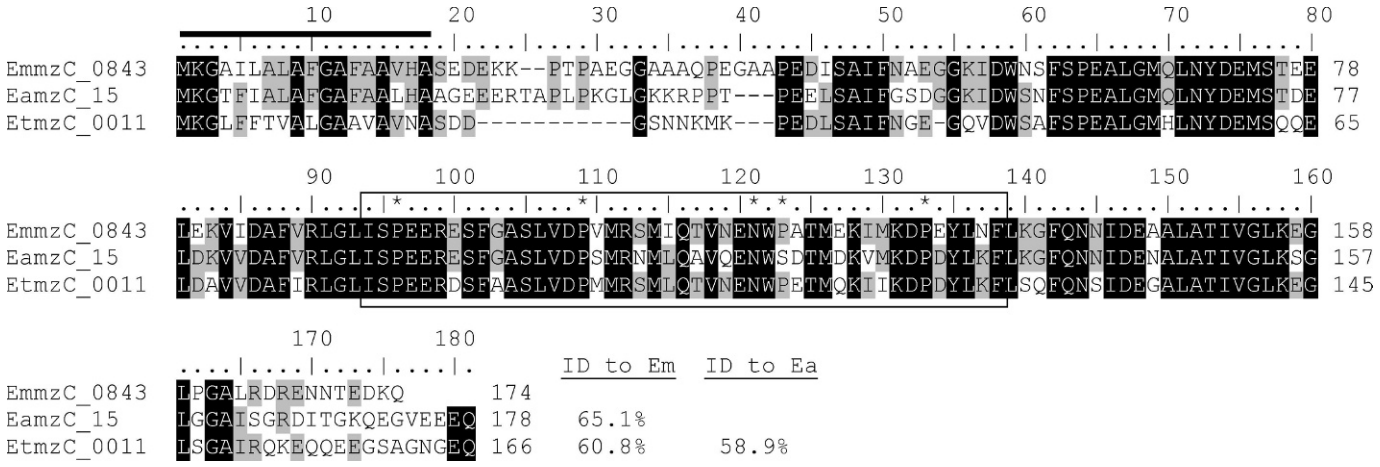


FIGURE 3. Alignment of the unknown *Eimeria* sp. specific predicted protein common to *Eimeria maxima* (Em), *Eimeria tenella* (Et), and *Eimeria acervulina* (Ea) merozoites. Predicted protein based on consensus sequences from EST clusters *Em*mzC_0843, *Et*mzC_0011, and *Ea*mzC_15. Alignment via MUSCLE (Edgar, 2004) with shading of identical (black) and 67% similar (gray) residues. The heavy bar indicates a predicted signal peptide region (aa 1–18), determined via SignalP 3.0 (Bendtsen et al., 2004). The boxed region covers a predicted STI1 heat shock chaperonin-binding motif (E = 1e-12) with highly conserved motif-specific residues indicated by asterisks. Percent interspecific sequence identity (ID) is given at the end of the sequences.

phenotypes such as drug resistance, precocious development, localization in the gut, and virulence. Finally, this EST database will be essential in providing cDNA evidence for the annotation of *Eimeria* genomes with gene models.

ACKNOWLEDGMENTS

We thank Ruth Barfield for technical assistance with merozoite collection and comments on the manuscript, and Alicia Beavers and Mónica Santín-Durán for generous assistance with sequencing. Support was also provided by the University of New Mexico's Molecular Biology Facility, which is supported in part by National Institutes of Health Grant P20 RR18754 from the Institute Development Award (IDeA) Program of the National Center for Research.

LITERATURE CITED

ADAMS, M. D., J. M. KELLEY, J. D. GOCAYNE, M. DUBNICK, M. H. POLYMEROPOULOS, H. XIAO, C. R. MERRIL, A. WU, B. OLDE, R. F. MORENO ET AL. 1991. Complementary DNA sequencing: Expressed sequence tags and human genome project. Science **252:** 1651–1656.

AJZENBERG, D., A. L. BAÑULS, C. SU, A. DUMETRE, M. DEMAR, B. CARME, AND M. L. DARDÉ. 2004. Genetic diversity, clonality and sexuality in *Toxoplasma gondii*. International Journal for Parasitology **34:** 1185–1196.

———, ———, M. TIBAYRENC, AND M. L. DARDÉ. 2002. Microsatellite analysis of *Toxoplasma gondii* shows considerable polymorphism structured into two main clonal groups. International Journal for Parasitology **32:** 27–38.

ALTSCHUL, S. F., W. GISH, W. MILLER, E. W. MYERS, AND D. J. LIPMAN. 1990. Basic local alignment search tool. Journal of Molecular Biology **215:** 403–410.

BARTA, J. R., B. A. COLES, M. L. SCHITO, M. A. FERNANDO, A. MARTIN, AND H. D. DANFORTH. 1998. Analysis of infraspecific variation among five strains of *Eimeria maxima* from North America. International Journal for Parasitology **28:** 485–492.

BECK, H. P., D. BLAKE, M. L. DARDÉ, I. FELGER, S. PEDRAZA-DÍAZ, J. REGIDOR-CERRILLO, M. GÓMEZ-BAUTISTA, L. M. ORTEGA-MORA, L. PUTIGNANI, B. SHIELS ET AL. 2009. Molecular approaches to diversity of populations of apicomplexan parasites. International Journal for Parasitology **39:** 175–189.

BENDTSEN, J. D., H. NIELSEN, G. VON HEIJNE, AND S. BRUNAK. 2004. Improved prediction of signal peptides: SignalP 3.0. Journal of Molecular Biology **340:** 783–795.

BLAKE, D. P., P. HESKETH, A. ARCHER, F. CARROLL, M. W. SHIRLEY, AND A. L. SMITH. 2005. The influence of immunizing dose size and schedule on immunity to subsequent challenge with antigenically distinct strains of *Eimeria maxima*. Avian Pathology **34:** 489–494.

CANTACESSI, C., S. RIDDELL, G. M. MORRIS, T. DORAN, W. G. WOODS, D. OTRANTO, AND R. B. GASSER. 2008. Genetic characterization of three unique operational taxonomic units of *Eimeria* from chickens in Australia based on nuclear spacer ribosomal DNA. Veterinary Parasitology **152:** 226–234.

CONWAY, D. P., AND M. E. MCKENZIE. 2007. Poultry coccidiosis: Diagnostic and testing procedures, 3rd ed. Blackwell Publishing, Ames, Iowa, 164 p.

DING, X., H. S. LILLEHOJ, R. A. DALLOUL, W. MIN, T. SATO, A. YASUDA, AND E. P. LILLEHOJ. 2005. *In ovo* vaccination with the *Eimeria tenella* EtMIC2 gene induces protective immunity against coccidiosis. Vaccine **23:** 3733–3740.

EDGAR, R. C. 2004. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research **32:** 1792–1797.

FERNANDEZ, S., A. C. COSTA, A. M. KATSUYAMA, B. N. MADEIRA, AND A. GRUBER. 2003. A survey of the inter- and intraspecific RAPD markers of *Eimeria* spp. of the domestic fowl and the development of reliable diagnostic tools. Parasitology Research **89:** 437–445.

FETTERER, R. H., AND R. C. BARFIELD. 2003. Characterization of a developmentally regulated oocyst protein from *Eimeria tenella*. Journal of Parasitology **89:** 553–564.

FITZ-COY, S. H. 1992. Antigenic variation among strains of *Eimeria maxima* and *E. tenella* of the chicken. Avian Diseases **36:** 40–43.

GENE ONTOLOGY CONSORTIUM. 2000. Gene ontology: Tool for the unification of biology. Nature Genetics **25:** 25–29.

HERM-GÖTZ, A., S. WEISS, R. STRATMANN, S. FUJITA-BECKER, C. RUFF, E. MEYHÖFER, T. SOLDATI, D. J. MANSTEIN, M. A. GEEVES, AND D. SOLDATI. 2002. *Toxoplasma gondii* myosin A and its light chain: A fast, single-headed, plus-end-directed motor. The EMBO Journal **21:** 2149–2158.

HETTMANN, C., A. HERM, A. GEITER, B. FRANK, E. SCHWARZ, T. SOLDATI, AND D. SOLDATI. 2000. A dibasic motif in the tail of a class XIV apicomplexan myosin is an essential determinant of plasma membrane localization. Molecular Biology of the Cell **11:** 1385–1400.

KUO, C. H., J. P. WARES, AND J. C. KISSINGER. 2008. The apicomplexan whole-genome phylogeny: An analysis of incongruence among gene trees. Molecular Biology and Evolution **25:** 2689–2698.

LEKUTIS, C., D. J. P. FERGUSON, M. E. GRIGG, M. CAMPS, AND J. C. BOOTHROYD. 2001. Surface antigens of *Toxoplasma gondii*: Variations on a theme. International Journal for Parasitology **31:** 1285–1292.

LETUNIC, I., R. R. COPLEY, B. PILS, S. PINKERT, J. SCHULTZ, AND P. BORK. 2006. SMART 5: Domains in the context of genomes and networks. Nucleic Acids Research **34:** D257–D260.

LEW, A. E., G. R. ANDERSON, C. M. MINCHIN, P. J. JESTON, AND W. K. JORGENSEN. 2003. Inter- and intra-strain variation and PCR detection of the internal transcribed spacer 1 (ITS-1) sequences of Australian isolates of *Eimeria* species from chickens. Veterinary Parasitology **112:** 33–50.

LI, G. Q., S. KANU, F. Y. XIANG, S. M. XIAO, L. ZHANG, H. W. CHEN, AND H. J. YE. 2004. Isolation and selection of ionophore-tolerant *Eimeria* precocious lines: *E. tenella*, *E. maxima* and *E. acervulina*. Veterinary Parasitology **119:** 261–276.

LI, L., B. P. BRUNK, J. C. KISSINGER, D. PAPE, K. TANG, R. H. COLE, J. MARTIN, T. WYLIE, M. DANTE, S. J. FOGARTY ET AL. 2003. Gene discovery in the Apicomplexa as revealed by EST sequencing and assembly of a comparative gene database. Genome Research **13:** 443–454.

MARRA, M. A., L. HILLIER, AND R. H. WATERSTON. 1998. Expressed sequence tags—ESTablishing bridges between genomes. Trends in Genetics **14:** 4–7.

MARTIN, A. G., H. D. DANFORTH, J. R. BARTA, AND M. A. FERNANDO. 1997. Analysis of immunological cross-protection and sensitivities to anticoccidial drugs among five geographical and temporal strains of *Eimeria maxima*. International Journal for Parasitology **27:** 527–533.

MISKA, K. B., R. H. FETTERER, AND R. C. BARFIELD. 2004. Analysis of transcripts expressed by *Eimeria tenella* oocysts using subtractive hybridization methods. Journal of Parasitology **90:** 1245–1252.

———, ———, AND G. H. ROSENBERG. 2008. Analysis of transcripts from intracellular stages of *Eimeria acervulina* using expressed sequence tags. Journal of Parasitology **94:** 462–466.

NG, S. T., M. S. JANGI, M. W. SHIRLEY, F. M. TOMLEY, AND K. L. WAN. 2002. Comparative EST analyses provide insights into gene expression in two asexual developmental stages of *Eimeria tenella*. Experimental Parasitology **101:** 168–173.

OPITZ, C., AND D. SOLDATI. 2002. "The glideosome": A dynamic complex powering gliding motion and host cell invasion by *Toxoplasma gondii*. Molecular Microbiology **45:** 597–604.

ROSE, E. M., AND P. L. LONG. 1962. Immunity to four species of *Eimeria* in fowls. Immunology **5:** 79–92.

SCHNITZLER, B. E., P. L. THEBO, J. G. MATTSSON, F. M. TOMLEY, AND M. W. SHIRLEY. 1998. Development of a diagnostic PCR assay for the detection and discrimination of four pathogenic *Eimeria* species of the chicken. Avian Pathology **27:** 490–497.

———, ———, F. M. TOMLEY, A. UGGLA, AND M. W. SHIRLEY. 1999. PCR identification of chicken *Eimeria*: A simplified read-out. Avian Pathology **28:** 89–93.

SCHWARZ, R. S., M. JENKINS, S. KLOPP, AND K. B. MISKA. 2009. Genomic analysis of *Eimeria* populations in relation to performance levels of broiler chicken farms in Arkansas and North Carolina, U.S.A. Journal of Parasitology **95:** 871–880.

SONG, X., L. XU, R. YAN, X. HUANG, M. A. A. SHAH, AND X. LI. 2009. The optimal immunization procedure of DNA vaccine pcDNA-TA4-IL-2 of *Eimeria tenella* and its cross-immunity to *Eimeria necatrix* and *Eimeria acervulina*. Veterinary Parasitology **159:** 30–36.

TABARÉS, E., D. FERGUSON, J. CLARK, P.-E. SOON, K. L. WAN, AND F. TOMLEY. 2004. *Eimeria tenella* sporozoites and merozoites differentially express glycosylphosphatidylinositol-anchored variant surface proteins. Molecular and Biochemical Parasitology **135:** 123–132.

TYZZER, E. E., H. THEILER, AND E. J. JONES. 1932. Coccidiosis in gallinaceous birds II. A comparative study of species of *Eimeria* in the chicken. American Journal of Hygiene **15:** 319–393.

VERCRUYSSE, J., D. P. KNOX, T. P.M. SCHETTERS, AND P. WILLADSEN. 2004. Veterinary parasitic vaccines: Pitfalls and future directions. Trends in Parasitology **20:** 488–492.

WAN, K. L., S. P. CHONG, S. T. NG, M. W. SHIRLEY, F. M. TOMLEY, AND M. S. JANGI. 1999. A survey of genes in *Eimeria tenella* merozoites by EST sequencing. International Journal for Parasitology **29:** 1885–1892.

XIE, M. Q., J. M. GILBERT, A. L. FULLER, AND L. R. MCDOUGALD. 1990. A new method for purification of *Eimeria tenella* merozoites. Parasitology Research **76:** 566–569.